Approved for Public Release

The NOAA Comprehensive Large Array-data Stewardship System (CLASS) Five-Year Plan

September 2002

Princeton Executive Square
5627 Allentown Road, Suite 100
Camp Springs, MD 20746

Table of Contents

1. Introduction

This document presents an overview of plans for the implementation and operation of the Comprehensive Large Array-data Stewardship System (CLASS) over the next five years. The plan focuses primarily on the archive and distribution functionality, since that is the most fully defined and highest priority component of CLASS at this time.

1.1 Overview of CLASS

The environmental stewardship mission of the National Environmental Satellite, Data, and Information Service (NESDIS) requires NESDIS to acquire, archive, and disseminate environmental data. NESDIS has been acquiring the data for more than 30 years, from a variety of in-situ and remote sensing observing systems throughout the National Oceanic and Atmospheric Administration (NOAA) and from a number of NOAA's partners. NESDIS foresees significant growth in both the data volume and the user population for the data, and has therefore initiated an effort to evolve current technologies to meet these future needs. This effort is CLASS.

The long-term goal for CLASS is the stewardship of all environmental data archived at the NOAA National Data Centers (NNDC). The initial objective for CLASS is to support specifically the following campaigns:

- NOAA and Department of Defense (DoD) Polar-orbiting Operational Environmental Satellites (POES)
- NOAA Geostationary-orbiting Operational Environmental Satellites (GOES)
- National Aeronautics and Space Administration (NASA) Earth Observing System (EOS)
- National Polar-orbiting Operational Environmental Satellite System (NPOESS)
- The NPOESS Preparatory Program (NPP)
- European Meteorological Operational Satellite (Metop) Program
- NOAA NEXt generation weather RADAR (NEXRAD) Program

CLASS will build upon systems already in place to implement an architecture for an integrated, national environmental data access and archive system to support a comprehensive data management strategy. The goals of CLASS are as follows:

1. Give any potential customer access to all NOAA (and some selected non-NOAA) data through a single portal.
2. Eliminate the need to continue creating "stovepipe" systems for each new type of data, while, as much as possible, using already polished portions/modules of existing legacy systems.
3. Describe a cost-effective architecture that can primarily handle Large Array-data sets, but is also capable of handling smaller data sets as well

The development of CLASS is expected to be a long-term, evolutionary process, as current and new campaigns are incorporated into the CLASS architecture. CLASS is intended to incorporate

and enhance the functionality of a number of existing NOAA systems, including the Satellite Active Archive (SAA) and NOAAServer. The major functions that CLASS is expected to perform include the following:

- Ingest of environmental data from CLASS data providers
- Extraction and recording of metadata describing the data stored in CLASS
- Archiving data
- Browse and search capability to assist users in finding data
- Distribution of CLASS data in response to user request
- Identification and location of environmental data that is not stored within CLASS
- Charging for data, as appropriate
- Operational support processes: 24x7 availability, disaster recovery

The development effort involves several different government and contractor groups in various geographic locations, affiliated with the Office of Satellite Data Processing and Distribution (OSDPD), the National Oceanic Data Center (NODC), the National Climatic Data Center (NCDC), and the National Geophysical Data Center (NGDC).

CLASS archive and distribution capabilities are operated and maintained by a government project manager and contractor employees under IPD's Central Satellite Data Processing (CSDP) contract with Computer Sciences Corporation (CSC). The government staff is responsible for managing the programmatic, technical, and administrative functions of CLASS, while the contractor staff is responsible for daily operations, adaptive maintenance, and development. The government and contractor personnel share responsibilities for system planning and for daily system monitoring and customer interaction.

CLASS will expand the SAA's archive and distribution support by providing duplicate operations at NCDC. NCDC, the permanent U. S. archive for POES data and derived data products, has supported the SAA in the past through a user-interactive Help Desk facility and through the provision of POES supporting documentation, including the NOAA Polar Orbiting Data (POD) User's Guide and the NOAA KLM User's Guide. Additionally, NCDC and SAA share data distribution responsibilities for Defense Meteorological Satellite Program (DMSP) data under a Memorandum of Understanding with the National Aeronautics and Space Administration (NASA) for the Earth Observing System (EOS) Program. In this upcoming fiscal year, full operations of CLASS will be provided both at the current OSDPD facility in Suitland, MD, and at the NCDC facility in Asheville, NC.

CLASS/SAA is currently supported by funding allocated through NOAA's Strategic Initiative for Seasonal-to-Interannual Climate Change. Additional resources are provided through the NOAA Earth System Data and Information Management (ESDIM) Program Office, the National Ice Center (NIC), and NOAA's Office of Research and Applications (ORA) for specific data management and data distribution services. Progress and performance are tracked through the NOAA-wide Information Technology (IT) initiative planning process by the NOAA Office of Administration as well as through the National Environmental Satellite, Data, and Information Service (NESDIS) Fiscal Year Operating Plan. NESDIS is the parent organization of OSDPD.

1.2 Background

The SAA, established as a demonstration prototype for electronic distribution of POES data in 1994, became operational in July 1995. During that first month, 379 Advanced Very High Resolution Radiometer (AVHRR) Level 1b data sets were distributed to 27 customers via the emerging Internet. In the seven years since that first operational distribution, average monthly volume has increased to nearly 120,000 data sets with a total size of 1.65 TB, and the SAA customer base stands at more than 20,000 active, registered customers. The active archive has been expanded during that period to include TIROS Operational Vertical Sounder (TOVS) data, Defense Meteorological Satellite Program (DMSP) data, Radarsat Synthetic Aperture Radar (SAR) imagery, NOAA CoastWatch products, operational (near-term) satellite-derived products, and climatic (time series) satellite-derived products. In addition, the SAA functions as an Information Technology (IT) platform for the distribution of several prototype satellite derived products, including NOAA's Geostationary-Orbiting Environmental Satellite (GOES) Sea Surface Temperature (SST) and Ocean Color (SeaWIFS) products.

Since 1995, enhancements to the SAA have also included:

- The development and refinement of the SAA World Wide Web (WWW) interface, including consolidation of three websites: the SAA main site, the SAA Products site, and the NOAA CoastWatch site (NCAAS)
- The automation of order and delivery services, subscription services, and bulk order processing
- The implementation of object-oriented software technologies

The SAA is implemented with a modern client-server architecture composed of clustered UNIX processors with access to the IPD Enterprise Server (IPD/ES), a robotic tape library, and several dedicated and broadband telecommunications hubs. The SAA provides data to customers in four basic modes:

- User-interactive access - Customer defines search criteria, views search results, and orders data through the SAA's World Wide Web (WWW) server
- User-interactive analysis - Customer accesses the products area of the web site and selects a climatic satellite derived product to be displayed. This product can then be downloaded using web browser capabilities.
- Subscription services - Customer defines criteria by which the SAA selects new data sets for automatic delivery
- Special bulk orders - Customer defines requirements for large amounts of data that the SAA makes available, generally through operator mediation, on a prearranged schedule.

All data are commonly distributed to customers through FTP services.

The SAA will now serve as the foundation for the archive, access, and distribution functionality for CLASS. Enhancements to the SAA will be required in order to support the new data providers and customers, including improved performance and availability as well as additional

functionality. At the same time, CLASS/SAA must continue to support the operational needs of the current customer base without interruption.

The CLASS development team is committed to establishing an operational environment that permits the infusion of new, improved access and distribution technologies with: 1) no negative impact on current customer satisfaction; 2) minimal impact on future operational funding; and 3) continuing improvement in the amount and quality of data and derived data products available through CLASS.

1.3 Current Status

CLASS is a fully operational system within OSDPD/IPD. In FY 2002, CLASS electronically distributed more than 20 Terabytes of polar satellite data, SAR data, and derived data products to its customers. CLASS has more than 4.9 million data sets on-line, including 83% of all NOAA AVHRR and TOVS data and DMSP data and a 100% of all NOAA CoastWatch data. The SAA's Historic Information Processing (HIP) effort is populating the CLASS near-line robotic library with NOAA and DMSP data spanning calendar years 1978 to CY 1994. At the completion of this effort, all of these data will be available on-line through CLASS.

In FY 2002, CLASS/SAA promoted two software releases (SAA Version 3, and CLASS Version 0), which introduced the following important software enhancements:

- Developed a new web-based user interface for IPD derived product data stored in CLASS. It is implemented with state-of-the-art technology, including a freeware off-the-shelf Web application framework, XML, Java servlets, and the Informix JDBC (JAVA Database Connectivity).
- Enhanced the CLASS Ingest system to handle many different kinds of IPD derived products. The enhanced system creates inventory records for product data files and converts data from native format to NetCDF (Network Common Data Form), a standard format that facilitates visualization of data in the SAA user interface. The new software achieves a high level of generality by using parameters specified in the database to define native file formats, NetCDF file attributes and variables, and the mapping of data from native formats to NetCDF.
- Developed a unified Web site that replaces the three SAA supported web sites (SWIG, LAS and NCAAS)
- Developed a Visualization Server to handle the generation of all images requested via the user interface. These include graphic displays of derived product data and low resolution browse images of POES orbital data
- Implemented XML message communication between the User interface and the Inventory Server and Visualization Server. Standard, off-the-shelf software is used to parse these XML messages. This new approach provides a uniform interface to all CLASS servers
- Extended the CLASS ingest process to include CoastWatch data, EDRMap data, and DMSP data in the new format used for satellite F16
- Designed a new inventory structure to replace the four separate inventories now in use for different types of data. The new inventory consists of one table containing common data

and several related tables, each holding data appropriate for a given data type. This new structure imposes uniformity and simplifies data access.

- Modified the Ingest process and the Inventory server to use the new inventory structure for CoastWatch data and planned the transition to the new inventory structure for all data types.
- Enhanced the product data ingest system to handle files with more complex record structures
- Enhanced the ingest system to accept data of varying quality. Instead of simply discarding data sets that fail some quality test, the system can now ingest those data sets, record the level of quality, and automatically replace them with data sets of better quality that are received later
- Developed a set of utilities to maintain the various caches that the Ingest and Delivery systems use for the temporary storage of data (e.g., ingested file and ordered files) and the permanent storage of data (e.g., browse image files and product data files). The utility functions include file storage and retrieval, file system cleanup, and resource locking to avoid conflicting access to files and file systems.
- Developed new functions that will be required to support distributed CLASS operations. These include: a checksum utility to be used in validating data copied between systems and recalled from archive, and a mechanism for maintaining duplicate browse image caches and other permanent caches at multiple operational sites.
- Provided support for ingesting and distribution of MSPPS data.
- Installed Secure FTP (SFTP) at request of the NCDC CLASS facility. The usage of SFTP was put forward as a pre-requisite by the NCDC facility for testing the ingest software at Asheville, NC.
- Made available to the user community Environmental Data Records Map (EDRMAP) products. The inclusion of EDRMAP products involved modifications to ingest and visualization software, and addition of product-related data to different database tables. The addition of EDRMAP products further generalizes the data ingest software and adds the capability of displaying polar-stereographic projections to the products web site.
- Prepared the system for processing NOAA-M data. This involved adding new satellite-specific parameters to various database tables and modifying code that extracts data for delivery.
- Extended subscription services to handle CoastWatch data.
- Complied all code with the new IBM compiler (VisualAge Version 5). Software improvements were made to bring code into compliance with the standards enforced by the compiler.
- Replaced various proprietary versions of container and string classes with the more portable versions of the Standard Template Library (STL) provided with the new compiler.
- Standardized all the makefiles and moved common macro definition rules to new shared make libraries to simplify maintenance of the build procedure.

Other accomplishments in FY2002 supported the transition from SAA to CLASS:

- Completed the detailed design of a new Delivery system to handle different types of orders (e.g., user, subscription, bulk) and to provide new functions for distributing IPD product data. The new design consolidates functions currently provided by separate systems that handle different kinds of orders and simplifies the database structure by consolidating several tables used to generate and track orders. It will provide users with many new options for ordering and receiving product data.
- Began working with NCDC to setup a duplicate CLASS environment at the NCDC facility in Asheville, NC.
    - Started coordination with NCDC personnel for installing a machine in Asheville and running the ingest software at Asheville.
    - Made the software and configuration changes needed to copy ingested files and browse images data files between sites and installed SAA ingest software at the NCDC CLASS facility in Asheville, NC.
- Constructed a new development environment that supports multiple CLASS development sites with a central software repository
- Replaced the Revision Control System (RCS) with the Concurrent Versions System (CVS) for managing source files. CVS is more suited to supporting remote-site development under CLASS
- Helped define the CLASS project goals for the next 18 months and delivered the project plan for the portion of CLASS work that it is under the Suitland Development team's control. This project plan served as a model for corresponding plans from all other CLASS teams.
- Defined CLASS risk management procedures and initiated risk identification and mitigation activities
- Developed and delivered the CLASS Configuration Management Plan, Quality Management Plan, and Software Development Guide. These documents establish the processes for use throughout CLASS and put the project on track for achieving CMM Level 2.
- In coordination with the CPMT, prepared and conducted a presentation to the CLASS oversight Group and to Greg Withee and Mary Glackin about the current status of the CLASS project and its planned activities for the next 18 months.
- Prepared and conducted a CLASS Developers workshop to provide a technical and process overview for the full distributed CLASS development team.
- Prepared and conducted a CLASS Earned Value Workshop to brief the CLASS teams on EV techniques and to define the specific procedures and tools for use on CLASS.
- Implemented nightly build process, to insure that software integration problems are identified and solved immediately.
- Delivered the CLASS Master Project Management Plan, which provides the overall project management guidance for all participating CLASS teams.
- Supported assessment and planning for introduction of e-commerce functionality into CLASS:
    - Initiated a time-boxed Business Area Architecture (BAA) to better understand the work-flows and business rules associated with the existing e-commerce, and order management processes. This process involved a series of interviews and a four-

day workshop held in Asheville, NC with representatives from each data center. The workshop was used to review, confirm and refine the e-commerce and order management processes as defined in the current system.
    - Delivered the e-commerce Buy-vs-Build analysis which indicates that buy is a better alternative
    - Helped coordinate with vendors, submitted questions that the COAST team had, coordinated site visits and provided guidance throughout the vendor selection process
- Consolidated CLASS requirements and wrote the Concept of Operations for the Data Archive, Access and Distribution Component of CLASS.
- Loaded the CLASS system requirements along with the CLASS Phase 1 requirements, the Goes Active Archive, the COAST, the NPP, the IJPS and the SAA requirements into the DOORS requirements management tool and built traceability matrices between the source requirements and the CLASS System Requirements.
- Updated the CLASS Software description.
- Updated the Data Dictionary and CLASS Database Entity-Relationship diagrams.

During FY2002, CLASS also completed a study of upgrades required to its hardware and operating systems. CLASS system administrators have begun implementation of the upgrade, including

- Configured new test environment by recycling two RS/6000 machines decommissioned from operations
- Relocated browse images from SCSI attached disks to a Network Attached Storage (NAS) device.
- Completed the CLASS hardware upgrade study for the Suitland facility. Identified candidates and key factors for use in product evaluation, worked with vendors to collect product information, and documented strengths and weaknesses of the candidate products. Visited different facilities with configurations similar to the one proposed in the study
- Prepared hardware installation and migration plan. The installation plan includes, installation schedule, blue prints of proposed equipment location, power and heat requirements and network requirements. The migration plan contains, the planned schedule for system migration, system test and parallel operations; along with data migration activities.
- Moved all data stored in the Network Attached Storage (NAS) device 760 to the NAS 840 and turned over the control of the NAS 760 to IPD System administrators
- Removed all X-stations used by software developers and operators and provided them with a suitable solution that does not affect productivity and does not decrease screen ‘real state’.

The CLASS project also participated in industry activities:

- Represented the Satellite Active Archive at the Coast Watch node managers meeting at Hawaii. Prepared the presentation for the conference and presented the current state of the NCAAS system, integration of NCAAS with CLASS and planned enhancements.

- Submitted abstract paper for the IEEE Mass Storage Conference

1.4 Document Organization

The system and technical enhancements proposed in this document are segregated into two general categories: near-term technological improvements that are scheduled to be implemented within the next twelve months; and longer-term improvements that should be implemented in the FY 2003 through FY 2006 era.

The short-term enhancements, described in Section 2, are dictated by: (1) new requirements for support from CLASS that was not required or available in the SAA; (2) known deficiencies in the current CLASS/SAA implementation, either through customer feedback or through simple "insider knowledge" by the CLASS staff; and (3) needed improvements rendered feasible through implementation of existing, proven IT technologies. Support for the upcoming EOS/MODIS, NPP, and Metop campaigns is the highest priority for short-term development efforts.

Longer-term enhancements, described in Section 3, may include improvements that fit near-term criteria, but that are impractical because of lack of time or funding. Most long-term enhancements, however, are based upon establishing an environment that: (1) can expand to handle the increased volumes and new types of data from new campaigns, platforms, and instruments; (2) can meet the new customer requirements for on-line access to high-volume, long time series data; and (3) can accommodate improvements and the infusion of new technology.

The short-term and long-term enhancements are each separated into the following four categories, which are considered critical:

- New Products & Services
- Customer Access
- Archive and Distribution
- System Maintainability

The discussions of near-term enhancements are more detailed in their statements of problems and their definitions of solutions, and probably more accurate in their estimates of costs. Costs associated with long-term enhancements should always be considered as estimates.

Tables that follow the sections on CLASS development include current, documented NOAA Performance Milestones; major, predicted events that will impact CLASS, such as satellite launches; and projected personnel loading tables for future CLASS management, operations, maintenance, and development staffs.

2. Near Term Enhancements (FY2003)

2.1 New Products and Services

Products and services offered by CLASS will be enhanced in the following areas:

2.1.1 Enhance CLASS to support GOES campaign

The CLASS team will define the requirements for archive and distribution support for GOES, and will then design and implement any changes necessary to CLASS to meet those requirements. CLASS is currently scheduled to be ready to support GOES by December 2003, so most of the development work will be completed during FY2003. The schedule will be revisited and refined after the CLASS team completes the requirements definition and analysis.

2.1.2 Enhance CLASS to support EOS/MODIS campaign

The CLASS team will work with NASA personnel to define the requirements for archive and distribution support for the EOS/MODIS campaign, and will then design and implement any changes necessary to CLASS to meet those requirements. CLASS is currently scheduled to be ready to support EOS/MODIS by December 2003, so most of the development work will be completed during FY2003. The schedule will be revisited and refined after the CLASS team completes the requirements definition and analysis.

2.1.3 Begin incorporating SABR functionality

In preparation for support for DMSP, the CLASS team will begin integrating the Satellite Archive Browse and Retrieval functionality into CLASS.

2.1.4 Complete requirements definition and analysis for NPP and Metop campaigns

Both the NPP and Metop campaigns are scheduled to begin delivering data to CLASS for archive and distribution in FY2005, with end-to-end tests planned for late 2004. During FY2003, the CLASS team will work with the NPP and Metop programs to define requirements for those campaigns and determine the changes needed to CLASS to support those requirements. Implementation of those changes is included in Section 3 under the long-term plan.

2.1.5 Security Enhancements

The CLASS team will assess the current software to ensure that secure FTP is supported for all users, and will work with the user community to expand the use of secure transfer of the data to customers.

2.1.6 Expand Access to Derived Products

This is an ongoing activity, continued from last fiscal year. The near-term enhancements discussed in the Archive and Distribution section of this plan (Section 2.3) will facilitate the addition of many new products over the next few years. (Projected numbers of new products

that will be added are shown in Section 4.3.) Eventually, CLASS will offer all IPD products, currently numbering around 200, along with GOES SST data, and PathFinder products.

2.1.7 Complete conversion to XML messages

During the last fiscal year, the CLASS project continued the conversion of the CLASS servers communication mechanism from a home grown messaging system to an industry standard mechanism, XML. Extensible Markup Language (XML) offers a simple and more straightforward approach for message interchange. CLASS servers are being modified to support messages in XML format and to provide responses in XML format. This enhancement will provide greater independence between servers and clients, as each one will process the XML message/response according to their needs. This conversion effort is expected to be completed early in FY2003.

2.2 Customer Access

The following improvements to the SAA Web interface are planned for the near-term.

2.2.1 Directory Services

The CLASS architecture will be modified to include an LDAP server to manage user information. This will provide CLASS with the capability to provide a central repository for user information, with the ability to share that information with other related systems, including the e-commerce system. The CLASS team will design the new functionality, acquire the necessary COTS components, and integrate the server with the other CLASS components.

2.2.2 Order Server

Continuing the progress towards providing web services, CLASS will encapsulate all Order input business logic into an Order Server that will receive XML formatted messages, will insert orders into the CLASS database, and return an XML formatted message with status information, including the recently created order number. This approach allows the client system to place orders for data delivery without needing to understand the database schema or CLASS business rules.

2.2.3 Database-driven Delivery Options

The new web interface has shown the benefits of having database driven search criteria. Adding new types of data for search is now very simple and requires minimal changes to the Web interface. The same approach will be used for the delivery options that are available for each data type. Having the available delivery options database driven greatly simplifies the process of adding new types of data, and reduces the Web Interface maintenance effort.

2.2.4 Improve Display of Search Results and Selection for Order

This is a continuing effort to improve the pages that display search results in order to help users select data sets. Navigation between results summary pages and results detail pages will be improved, and new options for sorting results will be provided. Options for easily selecting large numbers of data sets will be provided so that a user will no longer have to click on every data set

he wishes to order.

2.2.5 Improve data Visualization

This past fiscal year, CLASS added a visualization server to its pool of servers. This activity is a continuing effort to enhance visualization by allowing customer scaling of graphical depictions, and addition of new visualization tools.

2.2.6 Allow Access to Subscription Criteria

Allow subscribers access to their subscription criteria. Currently, CLASS operators manage all customer subscriptions. We will still retain control over the addition of new subscriptions while allowing customers to modify, delete, or add to their own subscription criteria.

2.3 Archive and Distribution

The following enhancements will significantly improve the storage and distribution of data.

2.3.1 Implement Geotiff Image Generation

Currently, CLASS generates GIF images, which are useful for browsing data but do not carry geographical information with them. By adding the capability to CLASS to produce Geotiff images, the images can be imported into other GIS systems for analysis and integration with other data. The CLASS team will enhance the visualization server to provide the capability to generate Geotiff images.

2.3.2 Improve Data Storage and Retrieval

In addition to the effort of replacing the robotic tape system (See 2.4.6), CLASS will analyze and prepare for storing large amounts of small data sets on disk, on what CLASS calls the permanent cache. Once the robotic tape system is replaced, CLASS will analyze all its capabilities and will upgrade the recall system to make better use of all its features.

2.3.3 Implement a Unified Catalog

During last fiscal year, CLASS developers designed a new inventory structure to replace the four separate inventories now in use for different types of data. The new inventory consists of one table containing common data and several related tables, each holding data appropriate for a given data type. The new structure imposes uniformity and simplifies data access. This new catalog database design is currently used for CoastWatch and Derived products. During this fiscal year we will migrate all other data types to this new unified catalog

2.3.4 Revise Data Type Naming Convention

The current CLASS data type hierarchy is not convenient for assigning parameters to data types. The CLASS team will define a more flexible naming convention for the data types that will better support control and management of the data.

2.3.5 Explore Geospatial Database Capabilities

The CLASS team will investigate the benefits of using Geospatial database capabilities provided by the CLASS DBMS. Implementation of such capabilities could enhance search capabilities, as well as other maintenance and performance characteristics of the system.

2.3.6 Upgrade SAA Network and Improve its Topology

In preparation for more and bigger data sets, CLASS has to evaluate its current network capabilities and topology and plan for upgrades. The following is a list of steps that the SAA is planning to take in order to improve its overall network performance:

- Upgrade all Ethernet cards to support 1 Gb and link them to the 1 Gb network.
- Directly link point-to-point the source of the data in the computer room with CLASS's Ingest machine. This will reduce network traffic and increase the overall network throughput.
- Move the CLASS network closer to the main Gateway and minimize its links with other networks.

2.3.7 Upgrade Product Data Search and Delivery

LAS enables a user to download selected product data in a variety of display formats, e.g., GIF images, spreadsheets, ASCII files. It does not, however, provide the product data in its native binary format or enable a user to select multiple data sets for subsequent batch delivery via FTP. The CLASS staff will implement both of these capabilities.

2.3.8 Integrate Bulk and Delivery Systems

With the implementation of the Bulk Recall and Bulk Order systems, some deficiencies of the current delivery system, especially when delivering large amount of files, were noticed. CLASS will incorporate the best of both approaches into two new systems: Recall and Delivery.

- The Recall System will be responsible for finding the requested data sets, in one of the multiple caches (Temporary, Visualization, or Permanent) or in the robotic system; and deliver it to the requesting process, usually the delivery system.
- The Delivery System will be responsible for processing the data set, formatting it according to user specifications, and delivering the formatted output to the user.

This approach will allow CLASS to efficiently handle different types of requests:

- Requests for small number of large data sets; service currently provided by the Delivery System.
- Requests for a large number of small data sets; service aimed to derived products data.
- Requests for large number of large data sets; service currently provided by the Bulk Order system.

2.4 System Maintainability

The CLASS staff will change software and procedures to make it easier to monitor and control operations and to add new capabilities to CLASS. The following areas have been identified:

2.4.1 Make Software Portable

All CLASS application C++ code will be reviewed to ensure that only C++ language constructs compliant with the ANSI standard are used and that only system functions compliant with the IEEE Portable Operating System Interface (POSIX) standard are invoked.

As a further step toward complete portability, CLASS personnel will explore the potential advantages of coding new processes in JAVA and converting existing processes to JAVA. The portability and standardization of JAVA, its avoidance of C++ features that foster coding errors (e.g., memory management, pointers, multiple inheritance), and its built-in utility and user-interface classes make JAVA a practical and appealing choice for an object-oriented programming language.

2.4.2 Enhance System Maintainability through Disk Sharing

This past fiscal year, the CLASS team assessed the feasibility of using Network Attached Storage (NAS) devices to facilitate sharing file systems among different machines in an efficient and secure way. The team determined that the NAS did not provide the capability that was needed, but instead will implement the SANergy shared file system as part of the new hardware architecture.

2.4.3 Improve Testing and Quality Assurance (QA)

This is an ongoing activity to improve the efficiency and effectiveness of system and regression testing through the use of new procedures, tools, and scripts. The CLASS test team will also ensure that testing includes the simulation of identifiable problematic situations that may be encountered in operations.

While release testing has been very successful in uncovering problems during system integration, the analysis of many of these problems reveals a need for more quality control in the earlier stages of design and coding. The CLASS team will implement new processes and procedures for peer review at each stage of development. CLASS has also initiated a quality management office to assist in process improvement.

2.4.4 Train Developers in New Technologies

CLASS is focused on maintaining currency with emerging information technologies that can help satisfy future customer requirements. The CLASS staff needs to be able to exploit these technologies to: 1) provide for the ingest, storage, and rapid distribution of an ever increasing volume and variety of data; 2) optimize system reliability, performance, and resource usage; and 3) ensure that the CLASS web interface continues to meet customers' needs and expectations and uses the latest proven technology available. To maintain currency with technology, the CLASS staff needs the training required to use new technologies advantageously.

2.4.5 Improve Operations Monitoring

The operations web page will be enhanced to report statistics about the number and nature of ingested data sets, display fast snapshot summary reports of the status of the system, report web

usage statistics, and allow the operator to query for log entries on the operational servers.

2.4.6 Complete Hardware Upgrade

This past fiscal year, the CLASS team completed an assessment of the hardware upgrade required to support planned operations. In FY2003, the team will install and configure the new hardware platform, and migrate the software and data to the new platform. The Upgrade path that CLASS is implementing is to move its processes out of the SP/2 complex to a set of Unix servers leaving the five SP/2 nodes for the Information Processing Division (IPD). This will be done because the most important benefit of having all CLASS machines as SP/2 nodes is the capability of sending a job to be executed in multiple nodes at the same time. The processes that CLASS executes benefit little or nothing from this high degree of parallelism, while many IPD processes can benefit from this.

CLASS will also move to a new robotics tape system and, once the migration is completed, CLASS will be completely decoupled from Enterprise Server, using it only as a source of data for ingest.

2.4.7 Install System at Second Site

In order to increase reliability and availability of CLASS, a second instance of the CLASS system will be installed at the NCDC facility in Asheville, NC. The CLASS team will complete installation of the system at that facility, installation of a T3 line to support the large volume data transfers, and testing of the system at that site. The CLASS team will also define and implement any software changes required to support the dual-site operations. The Suitland team will then conduct training for the new CLASS operations personnel in Asheville.

2.4.8 Process Improvement

As CLASS develops into a more complex system, handling increasing data volumes and customer requests, and increasing numbers of external interfaces, it becomes more important to ensure that the processes are in place for maximum efficiency and quality. The CLASS project will use the Software Engineering Institute's (SEI) 5-level Capability Maturity Model (CMM) to drive process improvement for the continued development of CLASS. The goal for FY2003 is to reach CMM Level 2, which represents industry best practices in project management, including Project Planning, Project Tracking, Requirements Management, Configuration Management, and Quality Management.

3. Long Term Enhancements (FY 2003 - FY 2006)

3.1 New Products and Services

3.1.1 Incorporate support for NPP and Metop campaigns

The NPP and Metop campaigns are scheduled to begin delivery of data to CLASS in FY2005. During FY2003, the CLASS team will work with those programs to define and analyze the requirements for those campaigns. The CLASS team will then implement any changes required to CLASS to support those requirements. Upgrades needed for these campaigns may require both hardware and software upgrades.

3.1.2 Incorporate DMSP support

The CLASS team will continue the short-term enhancement of integrating SABR functionality into the system. The long-term objective is for CLASS to provide all DMSP capabilities currently supported by SABR.

3.1.3 Incorporate Product Generation

For historic product data, (i.e., data not in the native binary archive), CLASS can provide an interface and automatic mechanism for batch reprocessing that will recreate the product data set from the CLASS standard archive. CLASS can also generate products as part of its Ingest processing.

3.1.4 Enhance NOAA Polar Pathfinder Reprocessing

The Pathfinder system encompasses the ingest of 1b data, the generation of various derived product data, and the distribution of both. CLASS can automate the generation of Pathfinder products and their installation in the CLASS derived product archive, where they will be available through the CLASS web interface. The following tasks are envisioned for enhancement to the NOAA Polar Pathfinder Reprocessing System architecture:

- Automate the creation of A1 data sets from 1C data sets.
- Ingest and catalog A1 data sets.
- Automate the creation of A2 data sets (daily, pentad, monthly)
- Ingest and Catalog A2 data sets.
- Convert A2 data sets to netCDF format for visualization

3.1.5 Provide Satellite Ephemeris Information

Enable a user to request equator crossing information and images of satellite ground tracks through the web interface.

3.1.6 Provide Rapid Implementation of Special Data Access Pages

The CLASS web customer interface will be enhanced to allow unique data access pages to be implemented within 24 hours of management request. These pages could be event-driven, such

as the 1998 implementation of the SAA "Brazil Fires Page," or be geographically oriented, such as a "Great Lakes Page." These pages will allow for predefined search criteria to filter ingested data for appropriateness. This implementation will include both 1b data and derived operational and climate data products.

3.1.7 Provide a Multimedia Presentation About CLASS

The CLASS team will create a presentation that can be shown to different audiences to educate them about CLASS and its services. This presentation will be useful for conferences or other meetings, where it is required to provide a professional looking introduction about CLASS.

3.2 Customer Access

Future plans for CLASS must allow for increasing demands from customers. CLASS must provide faster access to data, easier ways of browsing and ordering data, and the ability to access products derived from the data sets of interest. The following sections identify ways in which accessibility to CLASS data can be improved.

3.2.1 Allow Customization of User Interface

CLASS will offer the user customizable map manipulation features that will govern the generation and use of maps. CLASS will allow for additional columns and will offer sorting and ranging capabilities in the summary results pages.

3.2.2 Improve Geographic Search Capabilities

The enhanced geographic search algorithm has dramatically improved CLASS's ability to find data sets that overlap a user-defined search area. That initial implementation will be refined in several ways:

- Searches near the poles will be made more accurate.
- Users will be given the option to request only those data sets that overlap more than a user-specified percent of the search area.
- Users will be given the option to request only those data sets that overlap the search area by more than a user-specified minimum time interval.

3.2.3 Support Flexible Searches

CLASS will use fuzzy logic to allow the customer some flexibility with searches, allowing the customer to define a search with a text string containing keywords. The software will parse the string and extract the search criteria. This approach will allow a customer familiar with the CLASS parser to specify a query without having to traverse a variety of key-in fields and selection menus.

3.2.4 Create a Batch Search

Many customers are willing to wait for their search results as long as the results are accurate, reliable, and adequately fit their needs. Customers can identify cloud-free data sets by examining browse images, but this can be very time consuming. These customers would prefer that our search select only images that are cloud free, even if they have to wait for the search to be

completed. CLASS should offer customers an option to identify special attributes of the images they need. The resulting search will execute in very low priority batch mode.

3.2.5 Improve Mechanisms for Customer Interaction

In planning future enhancements, CLASS must take customer needs and expectations into account. Customers often volunteer their opinions via e-mail, but we can obtain a broader sampling of opinions with an optional customer survey form in which they could respond to specific questions regarding the selection and delivery of CLASS data. CLASS will provide a general form in the web interface and send special forms via e-mail to customers whose data needs could be better served by certain enhancements. CLASS will develop procedures and mechanisms for creating and updating survey forms, monitoring web interface usage, detecting patterns of usage that suggest the requirement for new options or capabilities, and distributing special survey forms to volunteer customers.

3.2.6 Support Interoperability with Other Data Distribution Sites/ Web Services

CLASS is currently undertaking a major development effort to expand its product offerings and improve product accessibility. As a result, CLASS will be well positioned to participate in the networked data access systems, virtual co-laboratories, and collaborative problem solving environments which are beginning to be developed in response to the availability of high speed networks (Next Generation Internet, Internet2, Abilene) and sophisticated middleware for connecting distributed databases. Examples of the use of such capabilities abound in the commercial sector. CLASS will readily participate in advanced, next-generation applications for advanced visualizations, virtual reality and data access, including next-generation, collaborative problem solving environments.

CLASS will be modified to convert all products to netCDF on ingest, and to store both the native and the netCDF formatted data, to provide a single standard data format for data sharing with users and other systems.

3.2.7 Provide Data Discovery Capability

CLASS will be enhanced to include the capability to locate environmental data that is stored in other NOAA archives, outside of CLASS itself. To support this, CLASS will incorporate and enhance the functionality currently provided by NOAAServer, utilize the FGDC standard for metadata, and work with providers to update available metadata.

3.2.8 Provide a Periodic Subscription

Currently all subscribers are subscribed to all available data sets. However there is a need for periodic subscriptions. CLASS will enhance the subscription system in such manner that a subscriber can specify:

- The time range when his or her subscription is valid, e.g. from May 1 to September 30.
- The lapse between data delivered, e.g. one data set every 24 hours.
- The time of collection, e.g. data sets collected between 12:00 and 20:00 hours UT

3.2.9 Provide an E-Commerce Capability

The CLASS team will define and implement the requirements to support recovery of the cost of providing data to customers. In this past fiscal year, CLASS supported the evaluation of COTS packages to handle the financial systems in support of e-commerce, and NOAA purchased the Oracle financial package. As the capabilities of the COTS package are evaluated, CLASS will define the appropriate billing and accounting interface and processes for CLASS customers.

3.3 Archive and Distribution

Speed of access to, and delivery of, data to our customers is paramount. Maintaining currency with telecommunication technologies and implementing the following proposed tasks will ensure continued support in these critical areas:

3.3.1 Perform Web Access Load Balancing

Once the new and enhanced web site is operational along with a large number of new derived data products, CLASS expects a large increase in the load of its web server. CLASS will analyze the usage patterns and will, where necessary, perform load balancing to ensure that all requests are served as quickly as possible.

3.3.2 Create Distributed Mirrored FTP Sites

The downloading of requested data sets could be made much faster through the establishment of FTP mirror sites, globally distributed in collaboration with other countries/ agencies.

3.3.3 Extract Data into Multiple Files

Currently, when a data set overlaps the user area more than once, the system will extract all overlapping scan lines and put then into a single file without making a distinction between the two overlap areas. The extract software will be modified to create multiple output files for different overlap areas at a user request.

3.3.4 Implement Geospatial Database Capabilities

Based on the results of the investigation into benefits of Geospatial database capabilities (see Section 2.3.5), the CLASS team will implement those capabilities that are determined to enhance the performance and functionality of the system.

3.4 System Maintainability

Maintaining the SAA system calls for constant improvements. Some ideas for improvements follow:

3.4.1 Develop GUI interface for Managing Database Tables

A new Graphical User Interface (GUI) tool will be developed that will enable operators to manage:

- Activity control tables, making it easier to add or change processing paths, process definitions, and trigger events.
- Ingest parameter tables, making simple the addition of new data types, new satellites or

new sources of data.
- Bulk Order tables, allowing for easy creation of bulk orders at user request.

3.4.2 Improve the Ingest Process

Improvements can be made in the Ingest subsystem to simplify operations and provide more checking of data quality. These improvements include implementing ephemeris quality checking (compare the data from equator crossing tables with the earth location information contained in data sets), and improved error reporting.

3.4.3 Consolidate Domain Constants

Values for operating parameters will be moved to a central repository in the database to ensure that all processes use the same values and that all modifications are done in a single location. This is an ongoing change that will increase the reliability and maintainability of the system.

3.4.4 Examine the benefits of Using New Capabilities of the RDBMS

The version of the RDBMS that CLASS is using allows for new data types to be defined and for C and Java functions to be added to the server. The usage of new data types can improve the design of the overall system, for example, by centralizing commonly used fields under one user-defined data type or by using collection data types to represent a list of values that should be treated as a single entity. The inclusion of data-intensive functions inside the server will increase performance by minimizing the transfer of data between the database server and its clients.

The benefits of using this new capabilities must be weighed against the cost of making the system less portable, as much of the new capabilities are only available for Informix, the RDBMS used by CLASS.

3.4.5 Upgrade CM tools

The current Configuration Management tools used by CLASS are adequate for its current needs but as new services are offered and the distributed CLASS team becomes more involved in development, the tools need to be reevaluated and upgraded to support the additional requirements. An additional source of change will be the IPD effort for standardization of all Configuration Management processes. CLASS will actively support those efforts and will upgrade its tools and procedures accordingly.

3.4.6 Plan for Hardware Upgrades

As new services and data are added to CLASS, usage will continue to increase. CLASS will start planning for the next technology upgrade needed to support these new services and data streams. Towards this end, CLASS has begun collecting and consolidating statistical data about the current systems usage. These data will continue to be collected and analyzed in order to decide when and what to upgrade.

3.4.7 Process Improvement

This is an ongoing effort for CLASS to continually improve the processes associated with

development and operations in order to improve efficiency and quality. The CLASS project will use the Software Engineering Institute's (SEI) 5-level Capability Maturity Model (CMM) to drive process improvement for the continued development of CLASS. The goal, at Level 5, is for all team members to be focused on continuous improvement, actively seeking out ways to improve the work processes and products.

4. Projected Schedules and Activities (FY 2003 - FY 2007)

4.1 Major Milestones

The continued development and enhancement of CLASS is primarily events-driven. These events may be either planned, such as platform launches, or unplanned, such as environmental events including hurricanes, typhoons, forest and grassland fires. The table below summarizes the known, planned events that impact the mission of CLASS. Data reprocessing and distribution caused by unplanned events generally have short-term milestones measured in days or hours. In FY 1998, for example, the SAA responded to a customer demand for high-resolution, polar-orbiting, satellite data over Brazil by rapidly developing specialized tools for selection, display, and delivery of that data in less than 24 hours.

CLASS is also driven by increasing customer demand for data, data products, and on-line services. To that extent, system enhancements to architecture, software, and processes are planned to meet expected demand. Such upgrades are included in the milestone chart.

CLASS Milestones	FY Goal
RADARSAT follow-on data on-line	FY 2003
DMSP F16 data on-line	FY 2003
GOES Implementation	FY 2003
CLASS Architecture	FY 2003
Add IPD Products on-line	FY 2002-FY 2004
Upgrade Operational Servers	FY2003
EOS/MODIS data on-line	FY2003
METOP-1 NOAA data on-line	FY2005
NOAA-N data on-line	FY2005
NPP data on-line	FY2005
NEXRAD data on-line	FY2006

4.2 Scheduled CLASS Releases (FY2003)

CLASS Version Number	Major Enhancements	Scheduled Release Date
1.0	• Single website to replace three combined • Unified catalog	Complete dev.: 12/02 Operational: 02/03
2.0	• Dual-site operations • GOES archive and distribution	Complete dev.: 06/03 Operational: 08/03

4.3 Performance Measures

The performance measures listed in the following table are indicative of metrics established through the NOAA Strategic Initiative for Seasonal-to-Interannual Climate Change.

CLASS Performance Measures	FY 2002	FY 2003	FY 2004	FY 2005	FY 2006	FY 2007
% U.S. POES archive on-line	85	95	100	100	100	100
% DMSP archive on-line	65	70	75	85	95	100
% GOES archive on-line		40	85	100	100	100
% EOS archive on-line		5	25	50	75	100
% Metop archive on-line				100	100	100
% NPP archive on-line				100	100	100
% NEXRAD archive on-line					10	40
% NPOESS archive on-line						0
Annual number of product data sets added	12	20	40	40	40	40
Monthly average data granules distributed (interactive orders)	70,500	78,000	85,000	90,000	95,000	100,000
Monthly average data granules distributed (subscription orders)	66,000	72,000	80,000	90,000	95,000	105,000
Monthly average data granules distributed (special bulk orders)	87,000	95,000	100,000	105,000	110,000	115,000
Monthly average data granules distributed (product data)	4,500	7,000	15,000	20,000	25,000	30,000
Total monthly average data granules distributed	228,000	252,000	280,000	305,000	325,000	350,000

4.4 Projected Staffing Levels

Area	FY2003	FY2004	FY2005	FY2006	FY2007
Technical Management*	4	4	4	4	4
Operations*	4	4	4	4	4
Maintenance	4	4	4	4	4
Development	15	17	17	14	14
Total	**27**	**29**	**29**	**26**	**26**

* Includes Government personnel